Exhibit A

UNITED STATES OF AMERICA
FEDERAL TRADE COMMISSION
WASHINGTON, D.C. 20580

Shira A. Steinberg
Bureau of Competition
(202) 326-3396
ssteinberg1@ftc.gov

July 16, 2024

BY ELECTRONIC MAIL

Katherine Trefz   Robert D. Keeling
Williams & Connolly LLP   Sidley Austin LLP
680 Maine Avenue SW   1501 K Street N.W.
Washington, DC 20024   Washington, DC 20005
ktrefz@wc.com   rkeeling@sidley.com

Re: *Federal Trade Commission, et al. v. Amazon.com, Inc.*, No. 2:23-cv-01495-JHC
(W.D. Wash.)

Dear Katie and Robert,

I write to follow up from and memorialize our July 3 and July 9, 2024 meet and confers regarding Amazon's proposal to collect, review, and produce documents using search terms and TAR 2.0 ("TAR").

Using Search Terms In Combination With TAR. Plaintiffs continue to have reservations about Amazon's proposed use of search terms in conjunction with TAR. As we noted in our letter, *see* Letter from S. Steinberg to R. Keeling at 1 (June 10, 2024) ("June 10 Letter"), and our call on July 3, if Amazon is going to use search terms in combination with TAR, it must do so in a manner that does not unduly narrow the population of documents that is ultimately reviewed and produced. Accordingly, as discussed further below, at a minimum, search terms and any proximity limiters used should be sufficiently broad to account for the use of TAR. Amazon's use of TAR means that some documents put into TAR review will not even be seen by human reviewers and introduces the possibility that responsive documents may be lost in the course of Amazon's review. Search terms introduce another avenue through which responsive documents may be lost, because even well-crafted search terms inevitably miss some responsive documents. Using search terms and TAR together compounds the risk that Amazon will fail to produce responsive documents. On the other hand, a properly calibrated TAR system will prioritize the documents that are more likely to be responsive for manual review and should

July 16, 2024
Page 2

eliminate the need to review documents that are highly unlikely to contain responsive material, after agreed-upon criteria have been met for the end of Amazon's review. As a result, the burden associated with broad search terms is much less than it would be if all documents captured by search terms were going to be manually reviewed. As we explained during our calls, we believe that the reasonable approach is to combine Amazon's use of TAR with a relatively broad set of search terms. Those search terms should sweep in substantially all documents that may be responsive to Plaintiffs' discovery requests; TAR will allow Amazon to prioritize its review to focus on the documents most likely to contain responsive material and cut off its review when it has reached an appropriate recall rate, and what remains is highly likely to be non-responsive.

You have proposed using separate search terms broken out for select custodians as opposed to using one set of search terms to collect documents from every custodian. *See* June 10 Letter at 2. However, as noted below, you were not able to fully explain the basis for Amazon's proposed four groups of custodians.

We are generally amenable to grouping (or "bucketing") custodians into a small number of groups, based on those custodians' responsibilities and the types of responsive information they are likely to have in their documents. We also agree with the approach of grouping search terms together into a small number of categories by subject matter. The search term and custodian proposals we are drafting will contain our proposed groupings. We appreciate your suggestion that a bucketing process may alleviate the need to negotiate search terms on an RFP-by-RFP basis and will make this process more efficient overall.

To be clear, all search terms should be applied to certain custodians—including the current and former CEOs, and current and former members of the S-team who have or had responsibilities relevant to the issues of Plaintiffs' Amended Complaint during their tenure as members of the S-team. Similarly, certain search terms—such as terms focused on the competitive process as a whole, or concerning Amazon's "flywheel" extending across both the Online Superstore and Online Marketplace Services markets—should be applied across all custodians.

TAR. During our calls on July 3 and 9, we discussed Amazon's June 27 TAR protocol proposal ("June 27 Proposed TAR Protocol") and raised a number of questions regarding Amazon's plans for its document review. We have integrated many of the points discussed on the calls in a modified proposed TAR protocol, attached herein at **Appendix A**, which includes amended provisions in line with our conversation on email threading, culling spam and junk documents prior to ingestion into Brainspace, and additional metrics Amazon should provide to Plaintiffs to allow for transparency regarding the review process. Plaintiffs reserve all rights to propose further edits or modify the proposal if the parties do not reach agreement or based on any additional information provided. We raised the following questions during the calls and now write to follow up with Amazon seeking additional information:

- You have represented "that Brainspace becomes unstable when applied to more than about 9 million documents." June 27, 2024 Letter from R. Keeling to S. Steinberg ("June 27 Letter") at 2. We asked for additional information on what issues you believe the model presents when over 9 million documents are ingested. You stated that an

July 16, 2024
Page 3

"unstable" model can still be used to review documents but that as more documents are ingested, it may take longer for the model to review and score all of them. You noted that sometimes the model can "error out" but this is fixable by restarting the model and letting it rescore documents. We asked for you to provide additional information on the errors the model may experience and more precisely, the maximum number of documents Brainspace can ingest before the model is no longer able to index documents. Please provide this information by **July 22** so that we can discuss ways for Amazon to address any capacity issues.

- Our understanding is that Brainspace *is* capable of working with more than 9 million documents, and that in any case the TAR models generated by Brainspace are portable, such that they can be applied to more than one dataset in parallel. *See Brainspace*, PORTABLE MODELS, https://brainspace-help.revealdata.com/en/Portable-Models.html. To the extent that Brainspace may in some cases be limited by compute capacity, that should not be an issue here given Amazon's extensive compute resources. As a result, we do not view this issue as one that should significantly constrain the scope of document review in this case.

- On July 9, you confirmed that Amazon planned to use email threading as part of its collection and review process. Plaintiffs want to ensure that Amazon's document population will not exclude lesser inclusive copies of email threads that contain unique family members, as required by the ESI Order. Dkt. #256 at § C(5). Plaintiffs also request confirmation that if Amazon withholds an email thread based on privilege, it will "produce any lesser inclusive copies of the email thread over which it does not claim the privilege or protection" Dkt. #256 at § E(4). To be clear, nothing in Appendix A is intended to relieve Amazon of its obligations related to email threading or privilege logging under the ESI Order. Please confirm by **July 22** that Amazon's use of email threading in conjunction with TAR will not conflict with the ESI Order.

- We raised concerns about using TAR to analyze standalone spreadsheets. We understand that spreadsheets, often containing large amount of numbers or data, are difficult for TAR models to analyze properly. You reported you had not experienced issues previously when using Brainspace to review spreadsheets, but we remain concerned based on our own experiences using TAR. On July 3, we asked for additional information on how Brainspace evaluates spreadsheets as part of training and its success in analyzing and properly ranking spreadsheets. Please provide this information for Plaintiffs' evaluation of Brainspace's capacity for document review in this action. In the interim, we have included spreadsheets as a filetype not suitable for TAR. *See* Appendix A at ¶ 4.1.

- We asked if Amazon would be able to provide "a metadata overlay with its productions listing the hyperlinks found in each produced email" pursuant to the ESI Order, *see* Dkt. #256 at 18, if it uses TAR in its review. Please identify any limitations that the use of TAR may impose on Amazon's ability to identify hyperlinks to internal document repositories and/or confirm that Amazon will still be able to produce such an overlay for documents reviewed by TAR.

July 16, 2024
Page 4

- During our call on July 9, we asked for additional information on what "quality control measures" Amazon plans on using "to identify certain documents within the Review Population that appear likely to have been coded incorrectly," to better understand how Amazon will confirm that its model is being trained properly. June 27 Proposed TAR Protocol at ¶5.2(c). You objected to providing this information on the grounds that these metrics constitute attorney work product; we disagree. The statistical measures used to evaluate a TAR model are not attorney work product and cannot be said to reflect an attorney's mental impressions or legal theories of a case. We asked whether Amazon was using industry-standard quality control measures; you reported that Amazon would be using non-standard measures that Sidley Austin has created on its own. It is concerning to learn that you plan to use non-standardized metrics for your quality control and object to providing further information to Plaintiffs. This is particularly troubling in light of the transparency and cooperation demanded of counsel when using TAR. *See, e.g.*, *In re Valsartan, Losartan, & Irbesartan Prods. Liab. Litig.*, 337 F.R.D. 610, 615 (D.N.J. 2020) ("[T]he backbone of TAR's use is transparency and collaboration…"). If Amazon stands on its objection and refuses to provide further information on the quality control measures it plans to use as part of its review, please provide any case law supporting your position so that Plaintiffs can evaluate your objection.

- As noted in our June 10 Letter and on our July 9 call, we remain apprehensive about Brainspace's ability to properly analyze documents that may be responsive to different theories of relevance in the case. In response, you pointed Plaintiffs to a glossary that appeared to confirm only that Brainspace has an algorithm for active learning called "Diverse Active," and you confirmed that Amazon plans to use that algorithm in its review. *See* Letter from R. Keeling to S. Steinberg at 5-6 (June 27, 2024). While we appreciate Amazon identifying the algorithm it plans on using, providing this name without additional information does little to address our concerns of the model's ability to account for multiple theories of relevance. The root of our concern is that our theories have overlap in Amazon's day-to-day business, and in turn its communications. It is still unclear how the model would evaluate a document that primarily focuses on Amazon's pricing conduct, and contains little text about the company's fulfillment practices against a document where the inverse is true. Both documents should be marked as responsive, but with Brainspace's textual analysis, one theory may be given greater weight than the other leading to inaccurate training of the model. On our July 9 call, you confirmed you are not using different tags for different theories, and believed the model could successfully identify documents as relevant, even if they related to different theories in the case. Please provide any information or documentation Brainspace or Consilio have to explain how Brainspace accounts for different theories of relevance in a case, particularly when there is likely to be significant overlap in documents pertaining to the different theories.

- Amazon's June 27 Proposed TAR Protocol proposes Plaintiffs accept a 75% recall rate in the abstract with no information about the document population itself or whether the recall rate will account for the use of search terms and TAR. As we have previously explained, if Amazon uses search terms to cull the population prior to using TAR, the recall rate Amazon proposes calculating will appear to reach 75% but in reality will be

July 16, 2024
Page 5

significantly lower because the method you propose does not account for the responsive documents left behind after the application of search terms. *See* June 10 Letter at 3. We previously proposed adopting a method of calculating recall that would account for documents being culled from the overall population prior to ingestion in the model. You did not respond to that proposal in your letter or directly on either of our calls. In the interest of reaching agreement on a TAR protocol without undue delay, we have proposed an alternative method of determining the recall rate based on relevant metrics that provide both parties with clarity into the population to make an informed decision about the appropriate recall rate. *See* Appendix A at ¶ 6.1. We believe this is a reasonable compromise, especially in light of the Western District of Washington's prior finding that a 75% recall target is "unacceptably low." *Garner v. Amazon.com, Inc*., 2023 WL 3568055, at *3 (W.D. Wash. May 19, 2023).[1]

We have two additional questions regarding Amazon's June 27 Proposed TAR Protocol that we did not have time to raise on the calls. We agreed to follow up with you in writing on these questions:

- First, we want to confirm the parties have the same understanding with respect to Amazon's assertion that it does "not intend to train the TAR model on documents responsive to Requests for which the parties have agreed that Amazon will respond on a 'go get' basis." While we agree that documents the parties agreed Amazon would collect and produce as "go gets" should not be coded as responsive for the purpose of model training simply because they were produced as "go gets," those documents should be coded as responsive to the extent they are captured by search terms and are responsive to non-"go get" document requests. In other words, "go get" documents should be treated the same as any other documents for the purposes of document review and model training. Please confirm that Amazon is in agreement.

- Second, in light of Amazon's representation in its June 27 Letter that you did not have "any repositories, custodians, or Requests" that you intend to review outside of TAR at this time, we propose that, prior to the parties finalizing the TAR protocol, Amazon either confirm this approach or disclose any custodians or repositories it plans on reviewing in a linear fashion outside of TAR.

To ensure we are making progress in line with the proposed milestones Plaintiffs set forth below, please respond to the questions above on or before **July 19, 2024.** We are available to

---

[1] We disagree with Amazon's attempt to distinguish the court's finding that a 75% recall target is "unacceptably low" in *Garner v. Amazon.com, Inc*., 2023 WL 3568055, at *3 (W.D. Wash. May 19, 2023). The court did not consider the efficacy of the search terms used in coming to its conclusion. Instead, it reasoned that "[t]he goal of discovery is the production of all responsive, relevant, non-privileged documents" and that "if 25% of responsive, relevant, non-privileged documents remain hidden" because the producing party is relying on TAR, "it may cost less, but it is no longer effective." *Id.* at n.5 at *3. We do, however, agree that because Amazon is planning, from the beginning, to use search terms in conjunction with TAR, the parties can account for the use of both by ensuring search terms are sufficiently broad and the TAR recall rate is calculated in a manner "to achieve a reasonable and proportional set of documents acceptable to both parties." June 27 Letter at 4.

July 16, 2024
Page 6

discuss any of our proposed edits to Amazon's June 27 Proposed TAR Protocol at your convenience.

Search Terms. As we explained on our July 3 call, Amazon has an obligation to work with Plaintiffs to craft search terms likely to result in relevant documents being collected, including by drawing on its institutional knowledge to propose terms that include project or code names and relevant language used in Amazon's ordinary course business documents. We discussed the parties exchanging search terms and asked Amazon to begin drafting its proposal. We expect Amazon has been working towards that goal, as Plaintiffs have, over the past two weeks. We propose the parties exchange initial search terms on **July 22.** Please confirm by **July 18** that Amazon agrees to that exchange date, or propose an alternative exchange date.

Custodians. During the call, we asked whether running separate sets of search terms for separate custodians would impact the TAR model. You confirmed the model can account for this and that it should not erroneously train the model. In your June 27 Letter, you proposed grouping Plaintiffs' RFPs into four "buckets" and groups of custodians to search for each bucket. However, your custodian proposal only identified custodians for approximately half of Plaintiffs' RFPs. During our July 3 call, we explained that your June 27 custodian proposal was missing employees – some of whom clearly are in possession of relevant information and that should be searched to respond to multiple RFPs. For instance, Andy Jassy, Amazon's CEO as of July 2021, was not included within your custodian proposal. We plan on circulating a counterproposal for custodians during the week of July 22.

We also asked about the rationale behind the four groups you divided our RFPs into for the purposes of your custodian proposal. You indicated that at least two buckets related to advertising and search results, respectively. We asked that you provide a description of each proposed category and your rationale behind those groupings. Amazon presumably had a rationale for its proposal, so should have this information readily at hand. Please provide it by **July 19**.

Document Collection. You confirmed your document collection is already underway. We requested Amazon collect documents dating back to 2014 through the present, rather than waiting to reach agreement on the period of collection for each discovery request in the case and having to repeat the collection process. As we explained, collecting documents twice – rather than collecting the full requested period at once – is inefficient and it remains unclear what impact, if any, this has to lessen Amazon's ultimate burden. Amazon claims there are two aspects to burden that are the basis of its refusal to collect documents from 2014 through the present: storage costs and cost of processing additional data. These costs are routine parts of discovery that nearly every party faces as part of the discovery process. Despite objecting to collecting documents on the basis of these two aspects of burden, Amazon has not quantified this burden over the course of months negotiating the scope of discovery (including the costs of collecting data while postponing processing) or provided an explanation as to why the burden outweighs the collection of highly relevant materials. We reiterate our request that Amazon now collect documents going back to 2014 through the present so that the progress of discovery is not slowed by the need for a later collection.

July 16, 2024
Page 7

Please also identify which custodians' files Amazon has collected to date, which collections are in progress, and please let us know by what date Amazon plans to have completed custodial collections for all custodians in its June 27 proposal. We expect that Amazon will provide information on a rolling basis about document collection and review so that the parties can meet and confer regarding the sequencing of custodians for the purposes of document collection, review, and production.

Non-Custodial Sources. The parties exchanged their ESI disclosures on June 20, 2024 which should have, amongst other things, identified "[a] list of non-custodial data sources (*e.g.*, shared drives, servers)" and a "list of third-party data sources" likely to contain discoverable ESI.[2] Dkt. #256. Those disclosures should have identified all sources of information relevant to the parties' claims or defenses, including but not limited to all information responsive to Plaintiffs' discovery requests. *See* Fed. R. Civ. P. 26(b)(1). Amazon must conduct a reasonable search for and of non-custodial sources, and Plaintiffs have frequently confirmed our understanding that Amazon will produce documents from non-custodial as well as custodial sources. *See, e.g.*, Letter from M. Baker to K. Trefz at 2 n. 2 (June 4, 2024). Amazon needs to disclose those non-custodial sources that may hold responsive documents so that the parties can meaningfully discuss Amazon's collection, review, and production of documents from those sources, and so that the review and production of those documents moves forward in parallel with the rest of Amazon's document production. However, Amazon's June 20 disclosure is seriously deficient.

First, rather than identifying any specific data sources in its June 20 ESI disclosures, Amazon merely referenced its February 22, 2024 letter, which does not come close to meeting the requirements set forth in the Court's June 11, 2024 order. It fails to identify which non-custodial sources are likely to contain discoverable information and the extent to which Amazon is able to preserve this information held by third parties, and actually *disclaims* that "[i]dentification of sources of information does not indicate that such sources or information are relevant or contain responsive information." *See* Letter from L. Mancari to E. Bolles at 3 (Feb. 22, 2024). Amazon must conduct a reasonable search for non-custodial and third-party sources of information relevant to the parties' claims or defenses, including but not limited to all information responsive to Plaintiffs' discovery requests. Please confirm by **July 22** that Amazon will do so and provide a deadline by which Amazon will provide an updated ESI disclosure. Alternately, state the basis for Amazon's refusal and any authority supporting Amazon's position.

Second, the parties have discussed several Amazon non-custodial sources that appear to clearly include discoverable ESI. Examples include ▮▮▮ *See, e.g.*, Amazon Responses & Objections to Plaintiffs' Second Set RFPs (objecting to the definition of "Collaborative Work Environment" and construing it as these five sources). This also includes ▮▮▮▮▮▮▮▮▮▮ a source Amazon agreed was an "internal document repository" from which Amazon would produce linked files. *See* Letter from

---

[2] Plaintiffs note that the ESI Order (Dkt. #256) requires Amazon to provide a further ESI disclosure for Plaintiffs' Fourth Set of Requests for Production within 10 days of serving its responses and objections to those requests. ESI Order, Section B.

July 16, 2024
Page 8

E. Bolles to C. Forkner at 1 (March 26, 2024). Amazon must search these non-custodial sources for responsive materials or explain why it cannot do so.

Third, there are other sources listed in Amazon's February 22 letter for which Plaintiffs need additional information to understand whether the source contains discoverable information and to negotiate Amazon's collection, review, and production of documents. One clear example is that Amazon has recently taken inconsistent positions in our correspondence about whether it will provide information about its use of Salesforce sufficient for Plaintiffs to assess whether Salesforce is a source of discoverable ESI. *See* Letter from K. Trefz to E. Bolles at 1 (July 12, 2024).

Fourth, Amazon's ESI disclosures fail to mention certain sources that Plaintiffs have learned through other means may contain discoverable ESI. ███████████████████ ██ ██ ██ ██ ██ ████████████████████████ Please explain why these repositories were not included in Amazon's February 22 letter or its June 20 ESI disclosures. These are precisely the types of non-custodial document sources that Amazon must identify and disclose, given Amazon's much greater access to information about its own documents and ESI.

Fifth, relatedly, during our July 3 call, we requested that you identify any non-custodial document repositories that Amazon plans to search using search terms. Please identify that subset of non-custodial sources by **July 19** and explain any limitations on how search terms are applied for those repositories.

Timing. As we discussed on our call, we believe the parties should work together to set milestones for document collection, review, and production to ensure that we are making progress in line with the Court's scheduling Order, Dkt. #159. In particular, Amazon should produce documents on a rolling basis and substantially complete its document production by the end of 2024 to ensure that the parties can move forward with depositions and conduct any necessary follow-up discovery in 2025. We propose the following timeline, which is generally based on the key dates we proposed during our July 3 call:

- **No later than July 22, 2024**: The parties have mutually exchanged initial search term proposals.

- **No later than July 31, 2024**: The parties have agreed on a TAR protocol and are in agreement on the majority of custodians and search terms.

- **No later than August 1, 2024**: As the parties continue negotiations regarding custodians and search terms, Amazon begins collecting documents for all custodians the parties are in agreement on, to the extent it has not already begun doing so. Amazon sets up its

July 16, 2024
Page 9

review platform and begins ingesting and processing documents. Amazon begins training reviewers and preparing to begin document review.

- **No later than August 22, 2024**: The parties have reached agreement on any outstanding custodians and search terms, or have reached an impasse on any disputes with respect to custodians, search terms, or other issues related to Amazon's document collection, review, and production. Any issues the parties have not resolved can be raised in the September Joint Status conference.

- **No later than September 3, 2024**: Amazon begins its document review and produces documents on a rolling basis, with productions every two weeks.

- **No later than December 20, 2024**: Amazon substantially completes its document production.

We believe this schedule will ensure that the parties have time to substantively engage on the issues while moving forward with discovery in a timely manner. Please let us know by **July 22** if Amazon agrees to these milestones or provide a counterproposal.

* * *

We are available to meet and confer regarding any of the topics discussed above.

Sincerely,

/s/ Shira A. Steinberg
Shira A. Steinberg
Attorney
Bureau of Competition
Federal Trade Commission